

**PEMBUATAN MODEL *DEPENDENCY PARSER*  
DAN *NAMED ENTITY RECOGNITION*  
MENGUNAKAN *LIBRARY SPACY* UNTUK  
BAHASA INDONESIA**

**Tugas Akhir**

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat  
Sarjana Teknik Informatika**



Dibuat Oleh :

**Daniel Chrisna Danuega**

**15 07 08258**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ATMA JAYA YOGYAKARTA  
2019**

## LEMBAR PENGESAHAN

**Pembuatan Model *Dependency Parser* dan *Named Entity Recognition*  
Menggunakan *Library Spacy* untuk Bahasa Indonesia**

Yogyakarta, 18 Juli 2019

**Daniel Chrisna Danuega**  
15 07 08258

Menyetujui,

Pembimbing I

Y. Sigit Purnomo WP., S.T., M.Kom.  
NPP: 04.01.710

Pembimbing II

Dra. Ernawati, M.T.  
NPP: 09.92.407

Penguji I

Y. Sigit Purnomo WP., S.T., M.Kom.  
NPP: 04.01.710

Penguji II

Yudi Dwiandiyanta B., S.T., M.T.  
NPP: 11.99.668

Penguji III

Findra Kartika Sari D., S.T., M.M.,  
M.T.  
NPP: 01.10.792

Mengetahui,  
Dekan Fakultas Teknologi Industri

Dr. A. Teguh Siswanto, M.Sc.  
NPP: 09.93.464

## PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Daniel Chrisna Danuega  
NPM : 15 07 08258  
Program Studi : Teknik Informatika  
Fakultas : Teknologi Industri  
Judul Penelitian : Pembuatan Model *Dependency Parser* dan *Named Entity Recognition* Menggunakan *Library Spacy* untuk Bahasa Indonesia.

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 28 Juni 2019

Yang menyatakan,



Daniel Chrisna Danuega

15 07 08258

## HALAMAN PERSEMBAHAN

*Remember that when you leave this earth, you can take  
with you nothing that you have received – only what  
you have given: a full heart, enriched by honest  
service, love, sacrifice, and courage*

*– Francis of Assisi –*

*Tugas Akhir ini saya persembahkan untuk*

*Tuhan Yesus Kristus,*

*Papah, Mamah*

*dan*

*Keluarga besar*

## KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena berkat dan bimbingan-Nya penulis dapat menyelesaikan laporan “Pembuatan Model *Dependency Parser* dan *Named Entity Recognition* Menggunakan *Library Spacy* untuk Bahasa Indonesia” ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana Teknik Informatika dari Program Studi Teknik Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis sadar betul bahwa dalam pembuatan laporan ini penulis banyak mendapat dukungan dan semangat serta bantuan yang sangat berarti untuk kelancaran penulisan dan pembuatan tugas akhir ini. Maka dari itu penulis ingin mengucapkan rasa terimakasih yang sebesar – besarnya kepada:

1. Tuhan Yesus Kristus yang hadir, memberikan berkat-Nya, pencerahan-Nya dan selalu membimbing penulis dalam penulisan laporan ini.
2. Papah Fransiskus Asisi Agus Susanto yang selalu mendoakan penulis dari atas sana, Mamah Veronica Indarningsih, Jimmy Nugraha Putra.
3. Keluarga besar Khouw yang selalu memberikan dukungan dan khususnya Ii Chen dan Ai Mei-Mei yang membantu perkuliahan penulis.
4. Bapak Dr. A. Teguh Siswantoro, selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
5. Bapak Y. Sigit Purnomo WP., S.T., M.Kom., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
6. Ibu Dra. Ernawati, M.T., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
7. Bapak Martinus Maslim, S.T., M.T., selaku Kepala Program Studi Teknik Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
8. Pak Supri yang membantu proses *remote* komputer.

9. Adel yang selalu mendampingi, memberikan semangat, dan menghibur disetiap proses penulisan laporan tugas akhir ini.
10. Teman – teman dari SMP Susteran Purwokerto yang menjadi teman mengerjakan tugas akhir bersama dan memberikan semangat satu dengan yang lainnya.
11. Sahabat PG (Bobby, Komang, Ian, Dika, Kevin, Travis, Wahyu, dan Verry) yang telah menjadi sahabat yang baik, menjadi tempat untuk berbagi, dan membangkitkan kembali semangat untuk menyelesaikan tugas akhir ini.
12. Grup bimbingan pak Sigit yang saling membantu dan berbagi info dalam proses pengerjaan tugas akhir.
13. Kepada semua pihak yang tidak dapat disebutkan yang juga berkenan membantu dalam kelancaran penulisan dan pengerjaan tugas akhir ini.

Demikian laporan tugas akhir ini dibuat, penulis paham bahwa dalam penulisan ini mungkin ada beberapa patah kata yang kurang berkenan atau tidak sesuai, penulis mohon maaf sebesar – besarnya. Semoga dapat bermanfaat bagi pembaca.

Yogyakarta, 28 Juni 2019



Daniel Chrisna Danuega

15 07 08258

# INTISARI

## PEMBUATAN MODEL *DEPENDENCY PARSER* DAN *NAMED ENTITY RECOGNITION* MENGGUNAKAN *LIBRARY SPACY* UNTUK BAHASA INDONESIA

Daniel Chrisna Danuega

15 07 08258

*Penggunaan teknologi NLP pada umumnya sudah mulai banyak digunakan dan khususnya di Indonesia sudah mulai digunakan di beberapa perusahaan. Salah satu teknologi yang digunakan untuk membuat model NLP yaitu library Spacy. Di dalam library Spacy khususnya untuk Bahasa Indonesia belum memiliki skema yang lengkap untuk digunakan sebagai model, sehingga pembuatan model Dependency Parser dan Named Entity Recognition dapat menjadi solusinya.*

*Model ini dibuat dengan cara pengkodean menggunakan bahasa pemrograman Python versi 3.x yang dibuat sesuai dengan aturan di library Spacy. Pembuatan model tersebut juga mengandalkan library Spacy dalam proses pemotongan kata (Tokenization), serta proses – proses NLP lainnya. Model ini akan menggunakan dataset dari Universal Dependencies sebagai dataset untuk pembuatan model Dependency Parser dan Named Entity Recognition. Metode yang akan dipakai merupakan metode yang sudah dibuat oleh Spacy di bahasa lain yang sekarang sudah bisa digunakan di Spacy. Model ini akan mengikuti metode dan alur algoritma yang sudah ada di library Spacy.*

*Hasil akhir dari pembuatan model ini akan berupa model Dependency Parser dan Named Entity Recognition yang siap digunakan untuk pemrosesan bahasa alami dengan Bahasa Indonesia. Nantinya model ini akan diunggah ke repositori Spacy untuk dapat dipergunakan secara luas.*

*Kata Kunci: NLP, Spacy, Dependency Parser, Named Entity Recognition, Bahasa Indonesia*

Dosen Pembimbing I : Y. Sigit Purnomo WP., S.T., M.Kom.

Dosen Pembimbing II : Dra. Ernawati, M.T.

Jadwal Sidang Tugas Akhir : 16 Juli 2019

## DAFTAR ISI

Judul.....	i
Lembar Pengesahan .....	ii
Pernyataan Orisinalitas & Publikasi Ilmiah .....	iii
Halaman Persembahan .....	iv
Kata Pengantar .....	v
Intisari .....	vii
Daftar isi.....	viii
Daftar Gambar.....	x
Daftar Tabel .....	xi
Daftar Rumus .....	xii
Bab I Pendahuluan .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian .....	3
1.5. Sistematika Penulisan.....	3
Bab II Tinjauan Pustaka .....	5
2.1. Tinjauan Pustaka .....	5
2.2. Landasan Teori.....	7
2.2.1. Natural Language Processing.....	7
2.2.2. Arsitektur Spacy.....	8
2.2.3. Dependency Parser.....	9
2.2.4. Named Entity Recognition.....	10
Bab III Metodologi Penelitian.....	12
3.1. Sumber Data.....	12
3.2. Alat Penelitian.....	12
3.2.1. Perangkat Keras .....	12
3.2.2. Perangkat Lunak.....	12
3.3. Langkah-Langkah Penelitian .....	12
3.3.1. Studi Pustaka.....	13
3.3.2. Pengumpulan Bahan.....	14
3.3.3. Pembuatan Dataset .....	14
3.3.4. Pembuatan Model dan Training Model.....	14



3.3.5. Evaluasi Model.....	15
3.3.6. Pembuatan Web Service .....	16
Bab IV Analisis dan Pembahasan .....	17
4.1. Pembuatan Dataset Dependency Parser .....	18
4.2. Training Model Dependency Parser.....	36
4.3. Pembuatan Dataset Named Entity Recognition .....	41
4.4. Training Model Named Entity Recognition.....	44
4.5. Penyatuan Model Parser dan NER.....	48
4.6. Evaluasi Model.....	49
Bab V Implementasi Model pada Web Service .....	51
Bab VI Penutup .....	55
6.1 Simpulan .....	55
6.2 Saran.....	55
Daftar Pustaka .....	56
Lampiran .....	58
1. Script CLI Spacy Convert conllu ke JSON (conllu2json.py) .....	58
2. Script CLI Spacy Evaluate (evaluate.py) .....	61
3. Script untuk Menghitung Skor Evaluasi (scorer.py).....	62
4. Script CLI Spacy Train (train.py) .....	65
5. Script Convert conllu ke JSONL (conllu2jsonl.py) .....	74
6. Script untuk Melatih Model NER (train-ner.py).....	75

## DAFTAR GAMBAR

Gambar 2.1. Arsitektur Spacy.....	9
Gambar 2.2. Tabel Daftar Label untuk Dependency Parser.....	10
Gambar 2.3. Visualisasi anotasi Dependency Parser.....	10
Gambar 2.4. Contoh Tabel Kelas Entitas.....	11
Gambar 3.1 Diagram langkah – langkah penelitian.....	13
Gambar 4.1. Salah satu teks kalimat beranotasi label relasi.....	19
Gambar 4.2. Bagian kode yang ditambahkan di dalam script spacy convert...	19
Gambar 4.3. Salah satu teks kalimat beserta anotasinya berformat JSON.....	20
Gambar 4.4. Algoritma Spacy dalam menentukan model terbaik.....	37
Gambar 4.5. Kode untuk menghentikan proses training (early stopping).....	37
Gambar 4.6. Algoritma untuk early stopping pada Spacy.....	38
Gambar 4.7. Struktur pohon direktori / folder model parser (id_ud-tag-dep-ner-1.0.0).....	41
Gambar 4.8. UI Prodigy untuk proses pengannotasian korpus secara manual...	43
Gambar 4.9. Struktur pohon direktori / folder model ner (model106).....	48
Gambar 4.10. Struktur pohon direktori / folder model parser dan ner (id_ud-tag-dep-ner-1.0.0).....	49
Gambar 5.1. Kode untuk mengakses model melalui web service.....	52
Gambar 5.2. Hasil response untuk dependency parser.....	53
Gambar 5.3. Hasil response untuk named entity recognition.....	53
Gambar 5.4. Contoh penggunaan web service untuk memprediksi named entity recognition.....	54

## DAFTAR TABEL

Tabel 2.1. Tabel pembandingan penelitian terdahulu.....	6
Tabel 4.1. Daftar skema universal dependencies labels.....	21
Tabel 4.2. Hasil epochs 1000 kali, dropout 0.2 dan batch size 1 hingga 64.....	39
Tabel 4.3. Hasil epochs 1000 kali, dropout 0.5 – 0.2 dan batch size 1 - 100.....	39
Tabel 4.4. Daftar kategori entitas pada skema OntoNotes.....	42
Tabel 4.5. Hasil losses pada pelatihan model NER.....	45

## **DAFTAR RUMUS**

Rumus 4.1. Rumus Akurasi.....	50
-------------------------------	----